

**統計学**

北門 利英（海洋生物資源学科）

Lecture 6 正規分布とその性質

Attention: 授業に関してわからないことがあれば、メールか6/17のリアルタイム接続時（13:00-14:00）に遠慮なく質問してください。

**Point:** 次の用語をしっかりと理解すること。

- 離散型確率変数と連続型確率変数の違い
- 正規分布の定義とその性質、そして確率の計算ができること

T. Kitakado      Statistics      Lecture 6 June 17, 2020

**背景例題**

T. Kitakado      Statistics      Lecture 06, June 17, 2020

生物資源学科の北門です。このビデオは統計学のLecture6の要約版です。今回は、正規分布の定義と性質についてお話しします。Lecture7以降でも正規分布は中心的な役割を果たしますので、必ず理解してください。それでは始めます。

## 例題

### 例題 1: 魚の体長の母集団の推測

養殖場で商用に飼育している魚種があり、体長 20cm 以上が商品サイズとして考えられているとする、飼育場の母集団の体長分布を調べるとともに、20cm 以上の個体数の割合も知りたいとする。どうすればそれらを知ることができるだろうか？ただし、魚はすべて同じ年齢とする。

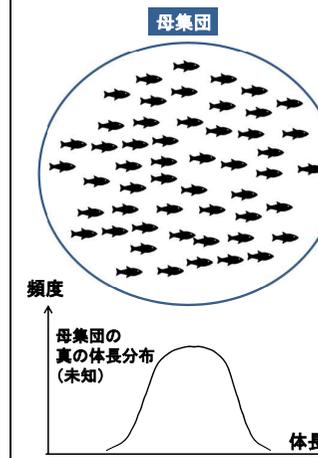
T. Kitakado

Statistics

Lecture 06, June 17, 2020

ここでは魚の体長の母集団について考えたいと思います。いま仮想的な設定として、養殖場で商用に飼育している魚種があり、体長分布と、母集団中の20cm以上のさなかの割合を知りたいとします。

## 母集団とサンプル集団（観測値）

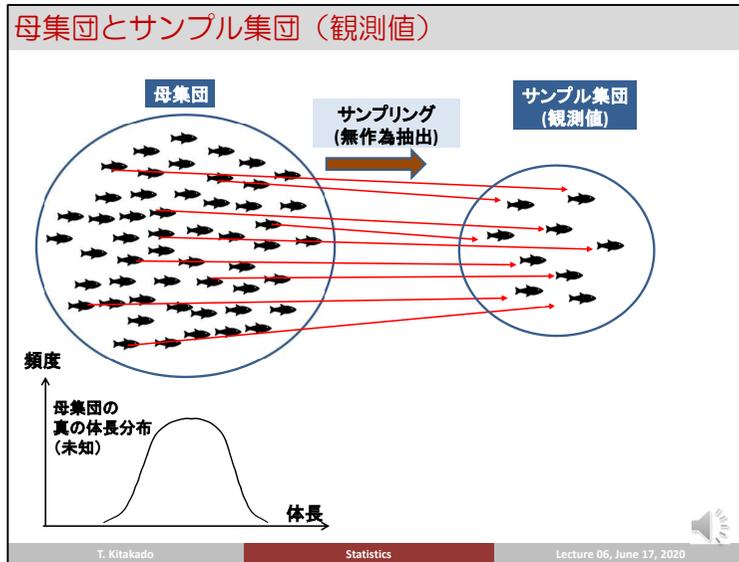


T. Kitakado

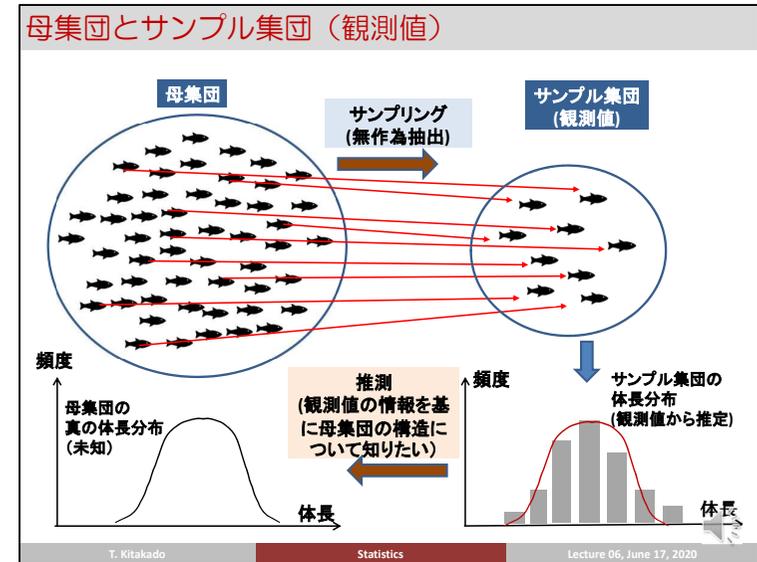
Statistics

Lecture 06, June 17, 2020

これを模式的に示すと、すべての魚の母集団があって、その体長分布を調べたいが、すべての個体の体長を測定するのは大変効率が悪いので...



一部を調べて全体の様子を知るために、サンプリングを行うのが常套手段です...



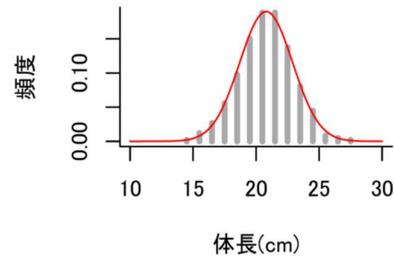
そしてサンプル集団の体長分布から、母集団の体長分布を表現したわけですが、そこになにかの数式を当てはめられないか、すなわち後で説明します正規分布を当てはめられないかと、考えます。

## 例題

### 例題 1: 魚の体長の母集団の推測

養殖場で商用に飼育している魚種があり、体長 20cm 以上が商品サイズとして考えられているとする。飼育場の母集団の体長分布を調べるとともに、20cm 以上の個体数の割合も知りたいとする。どうすればそれらを知ることができるだろうか？ただし、魚はすべて同じ年齢とする。

LenBin	LenClass	LenNumber
(10,11]	10.5	0
(11,12]	11.5	0
(12,13]	12.5	0
(13,14]	13.5	0
(14,15]	14.5	2
(15,16]	15.5	12
(16,17]	16.5	27
(17,18]	17.5	56
(18,19]	18.5	98
(19,20]	19.5	151
(20,21]	20.5	189
(21,22]	21.5	189
(22,23]	22.5	138
(23,24]	23.5	81
(24,25]	24.5	45
(25,26]	25.5	8
(26,27]	26.5	5
(27,28]	27.5	2
(28,29]	28.5	0
(29,30]	29.5	0



大泉ニジマス体長測定結果 (令和 1 年度) と正規分布の当てはめ結果

T. Kitakado

Statistics

Lecture 06, June 17, 2020

## 正規分布の定義

T. Kitakado

Statistics

Lecture 06, June 17, 2020

商用の例ではありませんが、大泉ステーションのニジマスの体長分布に当てはめると、データの度数分布表とヒストグラムがこのように得られ、結果としてこのような赤い線の曲線で、この体長分布を推定することができました。ここに正規分布の考え方を適用しています。

それではその正規分布の定義をしたいと思います。

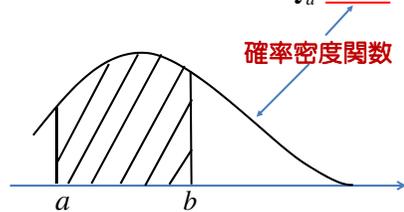
## 連続型の確率変数と確率分布

連続型確率変数：数直線上の値を連続的にとる確率変数

連続型確率分布：連続型確率変数に対する確率分布

標本空間 数直線のある区間

確率分布  $P(a < Y \leq b) = \int_a^b f(y)dy$



標本空間の特定の1点に対して確率を付与できない！

区間に対して積分で確率を定義する

T. Kitakado

Statistics

Lecture 06, June 17, 2020

その前に、正規分布も含まれる連続型確率分布について復習したいと思います。忘れた方はLecture2の資料も参考にしてほしいですが、端的に言いますと、2項分布やポアソン分布などの離散型とはことなり、原理的には数直線上のある区間のどの値でも連続的にとれる、という確率分布です。どの値が出やすいか、を表す関数を確率密度関数といいます。この図の場合、aよりもbの値の方が出やすいことは言えますが、横軸1点1点について確率を定義することはできず、代わりにこのように積分することで区間で確率を定義します。

## 連続型確率変数の期待値、分散、標準偏差

期待値(expectation)  $E[Y] = \int y f(y)dy$

分散(variance)  $V[Y] = E[(Y - E[Y])^2] = \int (y - E[Y])^2 f(y)dy$

標準偏差(standard deviation)  $SD[Y] = \sqrt{V[Y]}$

期待値と分散の性質 (離散型と同じ)

$$E[\omega_A \cdot Y_A + \omega_B \cdot Y_B] = \omega_A \cdot E[Y_A] + \omega_B \cdot E[Y_B]$$

$$V[\omega_A \cdot Y_A + \omega_B \cdot Y_B] = \omega_A^2 \cdot V[Y_A] + \omega_B^2 \cdot V[Y_B] \quad (Y_A \perp Y_B)$$

T. Kitakado

Statistics

Lecture 06, June 17, 2020

期待値、分散、標準偏差なども、シグマ記号が積分記号に代わるだけで、離散型と同じように定義できます。

## 正規分布の定義

連続型分布 1 確率変数  $Y$  の確率密度関数が

$$f(y) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad -\infty < y < \infty \quad \left( \begin{array}{l} -\infty < \mu < \infty \\ 0 < \sigma < \infty \end{array} \right)$$

となるとき,  $Y$  は正規分布 (normal distribution)  $N(\mu, \sigma^2)$  にしたがうという。

$$\int_{-\infty}^{\infty} f(y) dy = 1$$

期待値 (平均値)  $E[Y] = \int_{-\infty}^{\infty} y f(y) dy = \mu$

分散  $V[Y] = \int_{-\infty}^{\infty} (y-\mu)^2 f(y) dy = \sigma^2$

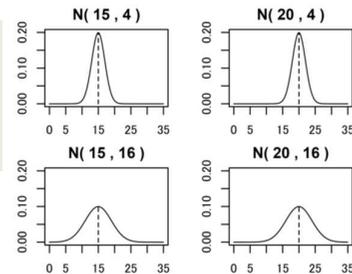


図4 正規分布の確率密度関数の概形

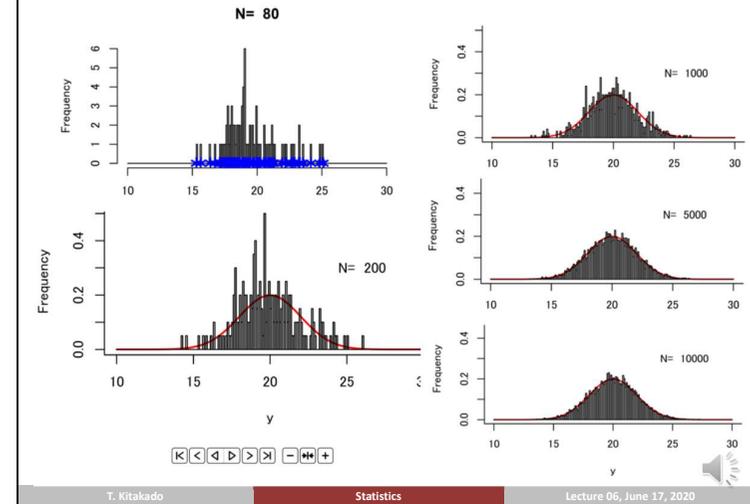
T. Kitakado

Statistics

Lecture 06, June 17, 2020

正規分布は連続型確率分布のなかでも最も使用頻度が高く, また様々な素晴らしい数学的性質を持っていますが, まずは確率密度関数がこのように表される確率分布だと思ってください。記号は  $N$ ,  $\mu$ ,  $\sigma^2$  乗, と書き,  $\mu$  は期待値,  $\sigma^2$  は分散を意味します。概形も平均値に対して左右対称で, 分散が一定なら確率分布の関数形は変わらずスライドするだけ, また分散が大きくなるとチャバリが大きくなり左右に裾を引く形になります。

## 確率変数の実現値と確率密度関数の関係



T. Kitakado

Statistics

Lecture 06, June 17, 2020

確率変数の実現値と確率密度関数の関係が分かりにくいかもしれませんが, ハンドアウトには  $N(20, 2^2)$  の実現値 200 個を順番に得てそれをヒストグラムに一つ一つ追加していく様子をアニメーションと, 200 個ずつ追加し実現値のヒストグラムと確率密度関数の関係の比較するアニメーションを用意しています。この図のように  $N=10000$  ではもう実現値のヒストグラムと確率密度関数はほぼぴったり一致していますね。

## 確率変数の実現値と確率密度関数の関係

### TUMSAT Statistics

Lecture 6: Normal distribution

Toshihide Kitakado

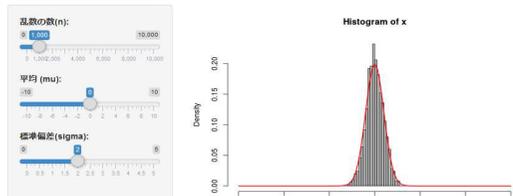
June 17, 2020



[https://kitakado.shinyapps.io/Lecture06\\_S2/](https://kitakado.shinyapps.io/Lecture06_S2/)

- ・正規分布  $N(\mu, \sigma^2)$  の概形と正規乱数ヒストグラムの重ね合わせ図です。スライダーを動かして、パラメータの意味をつかんでください
- ・乱数の数を増やすと、ヒストグラムが正規分布の確率密度関数の形にしっかり近づくことが分かるかと思えます
- ・サーバーのアクセス時間に制限がありますので、一人の利用は10分以内をお願いします。

#### 正規分布のパラメータと乱数の数



T. Kitakado

Statistics

Lecture 06, June 17, 2020

## 正規分布の基本性質

T. Kitakado

Statistics

Lecture 06, June 17, 2020

なお、このスライドの記載のスライドに、インタラクティブなアプリを用意しています。ぜひ覗いてみてください。

では次に正規分布の基本性質です。2つ述べますが、今回は最初に述べる標準正規分布をしっかり頭に入れてください。

## 正規分布に関する性質 (1) 標準正規分布 $N(0,1)$

性質 1 確率変数  $Y$  が  $N(\mu, \sigma^2)$  にしたがりとき,

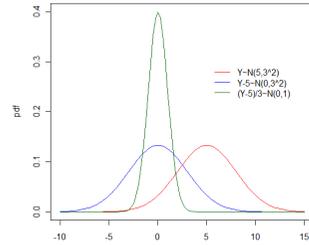
$$Z = \frac{Y - \mu}{\sigma} \quad (5)$$

は  $N(0, 1)$  にしたがり, この変換を標準化 (standardization), また  $N(0, 1)$  を標準正規分布 (standard normal distribution) という.

$$Y \sim N(\mu, \sigma^2)$$

$$Y - \mu \sim N(0, \sigma^2)$$

$$\frac{Y - \mu}{\sigma} \sim N(0, 1)$$



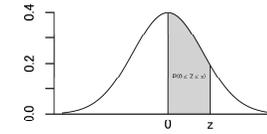
T. Kitakado

Statistics

Lecture 06, June 17, 2020

この性質はとても重要で, 定期試験でも必ず出ますが, まずある確率変数  $Y$  が平均  $\mu$ , 分散  $\sigma^2$  の正規分布に従っているとします. このとき, この  $Y$  から平均値  $\mu$  を引いて, 標準偏差  $\sigma$  で割った値も確率変数ですが, それを  $Z$  と置くと, もととの  $\mu$  と  $\sigma$  が何であってもそれを使って変換すれば平均が  $0$ , 分散が  $1$  の正規分布に変換できます. このような変換を標準化, また平均  $0$ , 分散  $1$  の正規分布を標準正規分布と呼びます.

## 正規分布に関する性質 (1) 標準正規分布 $N(0,1)$



$$Z \sim N(0,1)$$

$$P(0 \leq Z \leq 1.96) = 0.4750$$

	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817

T. Kitakado

Statistics

Lecture 06, June 17, 2020

この標準正規分布には数表が与えられていて, 例えば  $0$  から  $1.96$  の値の確率は  $0.475$  という具合に, この数表を用いた計算の練習問題をハンドアウトに記載していますので, ぜひ取り組んでください.

## 正規分布に関する性質 (2) 再生性

正規分布にしたがう独立な確率変数の和もまた正規分布にしたがう。独立同一な  $n$  個の正規分布にしたがう確率変数の和もまた正規分布にしたがう。これらの性質は区間推定と仮説検定のところで再度説明するので、今回は定理の存在だけ覚えておいてください。

定理 2 確率変数  $Y_1, Y_2$  が独立でそれぞれ  $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$  にしたがうとき、 $aY_1 + bY_2$  は  $N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + b^2\sigma_2^2)$  にしたがう。

定理 3 確率変数  $Y_1, Y_2, \dots, Y_n$  が独立同一に  $N(\mu, \sigma^2)$  にしたがうとき、 $\bar{Y} = \sum_{i=1}^n Y_i/n$  は  $N(\mu, \sigma^2/n)$  にしたがう。



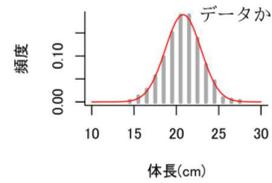
再生性については来週またお話しします、

## 例題と演習



続いて最初の大泉ニジマスの例に戻ると...

## 例題の解



データから平均値  $\mu = 20.8$ , 標準偏差  $\sigma = 2.1$  と推定される.

$Y \sim N(20.8, 2.1^2)$  を想定

推定の方法は来週詳しく述べます.  
今回は推定のごとは忘れ、正規分布の性質と確率の計算に集中します

図2 大泉ニジマス体長測定結果 (令和1年度) と正規分布の当てはめ結果

T. Kitakado

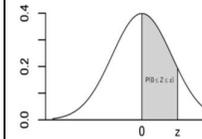
Statistics

Lecture 06, June 17, 2020

来週お話しする推定の方法を通して、 $\mu=20.1$ ,  $\sigma=2.1$ と推定されますので、これを基にニジマスが20cm以上である確率を計算すると...

## 例題の解

$$\begin{aligned}
 & P(Y \geq 20) \\
 = & P\left(\frac{Y-20.8}{2.1} \geq \frac{20-20.8}{2.1}\right) \quad (Y \sim N(20.8, 2.1^2)) \quad (\text{括弧の中は事象を表すので両辺四則演算可, 標準化する}) \\
 = & P(Z \geq -0.381) \quad (Z = \frac{Y-20.8}{2.1} \text{ とおくと } Z \sim N(0, 1)) \\
 = & P(-0.381 \leq Z \leq 0) + P(Z \geq 0) \quad (Z \sim N(0, 1) \text{ の数表 (付録) を用いるために確率を分割}) \\
 = & P(0 \leq Z \leq 0.381) + 0.5 \quad (Z \sim N(0, 1) \text{ の確率密度関数は } 0 \text{ に対して左右対称}) \\
 = & P(0 \leq Z \leq 0.38) + 0.5 \quad (P(0 \leq Z \leq 0.38) \text{ は付録の数表から } 0.148 \text{ ととる}) \\
 = & 0.648
 \end{aligned}$$



	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389

T. Kitakado

Statistics

Lecture 06, June 17, 2020

形式的にこのように定義でき、確率の中の事象を少し変形して標準正規分布の形に持っていきます。途中Zが0以上である確率が出てきますが、これは0で左右対称ですので0以上あるいは0以下の確率はともに0.5、そして0~0.38に入る確率は0.148ですから併せて0.648となります。簡単ですね。

## 練習問題

練習問題 1 ある魚の体長  $Y$ (cm) が正規分布  $N(20, 5^2)$  にしたがうとき、 $P(10 \leq Y \leq 30)$  および  $P(Y > 25)$  を求めよ。(答え:0.9544 と 0.1587)

練習問題 2 統計学の定期試験の点数が正規分布  $N(65, 4^2)$  に従うとする。60 点以上で合格とするとき、合格者の割合はいくらか? (フィクションです) (答え:0.8944)

練習問題 3 あるアザラン種は、資源量が 1000 頭以下である確率が 30 パーセント以上あるとき、絶滅危惧種とみなされるとする。本種に対する最新の調査報告書によると、資源量推定値の自然対数が確率分布  $N(7.2, 0.6^2)$  に従うと記載されていた。この調査の結果、この種は絶滅危惧種とみなされるか? (これもフィクションです)

クジラの資源量を  $N$  とおくと、調査の結果、 $Y = \log N \sim N(7.2, 0.6^2)$  が成り立つ。いま知りたい確率は  $P(N \leq 1000)$  であるから

$$\begin{aligned} P(N \leq 1000) &= P(\log N \leq \log 1000) = P(Y \leq \log 1000) = P\left(\frac{Y-7.2}{0.6} \leq \frac{\log 1000 - 7.2}{0.6}\right) \\ &= P(Z \leq -0.487) = P(Z \leq -0.49) = 0.3121 \end{aligned}$$

と計算され、したがって本種は絶滅危惧種とみなされることになる。

次の問題は提出課題です。次回以降の提出となりますので今回は提出不要です。

練習問題 4 あるチョコレート工場で 100g の板チョコを生産しているが、製品によってばらつきが生じ、正規分布  $N(102, 2^2)$  に従うとされている。100g 未満のチョコレートは出荷できないとき、生産したチョコレートの何%が不良品となるか? また、不良品率を 1 パーセント以下にしたいとき、板チョコの重さの平均値をいくらにするように生産工程を変えればよいか? (平均値を変えてもばらつき、すなわち分散は変わらないとする)

T. Kitakado

Statistics

Lecture 06, June 17, 2020

理解の確認のために練習問題1~4にも取り組んでみて下さい。練習問題4は次回以降とまとめてとなりますが、提出課題です。それではLecture6のダイジェストビデオを終わります。